

Hilgardia

A JOURNAL OF AGRICULTURAL SCIENCE PUBLISHED BY
THE CALIFORNIA AGRICULTURAL EXPERIMENT STATION

Volume 61 • Number 1 • November 1994



The Predictive Models and Procedures Used in the Forest Stand Generator (STAG)

**Greg S. Biging, Timothy A. Robards,
Eric C. Turnblom, and Paul C. Van Deusen**



ABSTRACT

The Forest Stand Generator, STAG, is a microcomputer-based program that uses statistical routines to produce a projection set composed of individual tree measurements of diameter at 4.5 feet above ground (called the diameter at breast height or DBH), total height, height-to-crown base, species, and tree expansion factor. When data sets are not complete, STAG can be used to produce a projection set for a wide class of inventory procedures. The authors will discuss the estimation procedures used by STAG to (1) fill in missing measurements of tree height, height-to-crown base, or both; (2) generate stands from summary statistics; and (3) convert stand table data—numbers of trees by DBH classes and species—to individual tree records, so that these projection sets, composed of complete individual tree records, can be analyzed by the California Conifer Timber Output Simulator (CACTOS) for simulation of tree growth and mortality, even though the initial data sets could not have been used with CACTOS. Also discussed are the predictive equations and analytic procedures used to produce a projection set for these three categories of data availability. Recommended uses of STAG and the type of data required for accurate development of projection sets are presented.

THE AUTHORS:

Greg Biging is Associate Professor, Department of Environmental Science, Policy and Management, University of California, Berkeley.

Timothy A. Robards is Sustained Yield Forester with the California Department of Forestry and Fire Protection.

Eric C. Turnblom is former Postgraduate Researcher, Department of Environmental Science, Policy and Management, University of California, Berkeley, and is now Assistant Professor at the University of Washington, Seattle.

Paul C. Van Deusen is Research Scientist at the USDA Forest Service Southern Forest Experiment Station, New Orleans, Louisiana.

Continued inside back cover

The Predictive Models and Procedures Used in the Forest Stand Generator (STAG)¹

INTRODUCTION

The interior forests of northern California are typically composed of mixed conifer species of multiple ages and sizes. Inventory procedures for these lands are varied, as is true in the rest of the United States. This work addresses three common inventory procedures. One typical procedure is to measure diameters at breast height (DBH) and to subsample tree heights (H) and heights-to-crown base (HCB). This procedure yields what can be considered a *missing data* case. Another common procedure is to record the number of trees by diameter classes. This yields *stand table data*, which are a discrete approximation of the continuous diameter distribution. In some cases only *stand summary statistics* are recorded, such as the basal area per acre (basal area is the cross-sectional area of trees measured at 4.5 feet above ground in square feet on a per acre basis) and number of trees per acre. For summary statistics no individual tree information is recorded, just overall stand parameters.

A common use of inventory data is to simulate the future growth and yield of the stands from which the data were derived. Our goal is to ensure that the three forms of inventory data listed above can be made to conform to the requirements of the California Conifer Timber Output Simulator, known as CACTOS (Wensel, Daugherty, and Meerschaert 1986; Wensel, Meerschaert, and Biging 1987; Wensel and Biging, 1987). CACTOS simulates the growth and development of individual trees and requires that species, DBH, H, HCB or live crown ratio, and tree expansion factor² be supplied for each individual tree that contributes to the stand description. When all these data are present, the stand description becomes a minimal projection set, which is composed of the individual tree records needed for growth simulation. To take full advantage of the simulation capacity of CACTOS, these variables should be measured for all trees.

When data sets are missing any of the components necessary for simulation purposes, the Forest Stand Generator, STAG,³ can be used to produce a projection set for a wide class of inventory procedures (Biging et al. 1995). This article discusses the estimation procedures used in STAG to (1) fill in missing measurements of tree height, height-to-crown base, or both; (2) generate stands from summary statistics; and (3) convert stand table data—numbers of trees by DBH classes and species—to individual tree records, so that these projection sets (composed of complete individual tree records) can be analyzed by CACTOS. We also discuss the predictive equations and analytic procedures used to produce a projection set for the three categories of data availability.

¹Accepted for publication October, 1993.

²The crown ratio is the proportion of tree bole length that occurs within the crown of the tree. The tree expansion factor is the number of trees per acre that the sample tree represents.

³STAG is an acronym for Forest STAnd Generator for mixed Conifer Species, copyright © Regents of the University of California, 1986-94.

DATA USED FOR MODEL DEVELOPMENT AND TESTING

Data for this study were provided by the Northern California Forest Yield Cooperative's Growth and Yield Project. These data were collected from 710 permanent plots located throughout the mixed conifer region of northern California. The variables measured for each tree included species, DBH, H, and HCB. The permanent plots were established in 1978-79 and a five-year re-measurement was made in 1983-84. These plots were typically $\frac{1}{2}$ acre in size and contained subplots used to measure submerchantable trees. Usually trees greater than 11.0 inches in DBH were measured on the full plot. Trees between 5.5 and 11.0 inches in DBH were measured on a $\frac{1}{10}$ acre subplot, and trees between 1.5 and 5.5 inches in DBH were measured on a $\frac{1}{20}$ acre subplot. There were some variations in the class limits, depending on the company collecting the data. The five-year re-measurement data were used for the models developed in later sections of this article. Figure 1 shows the location of the permanent plots by township and Appendices A and B provide summary statistics for much of the data used in this study.

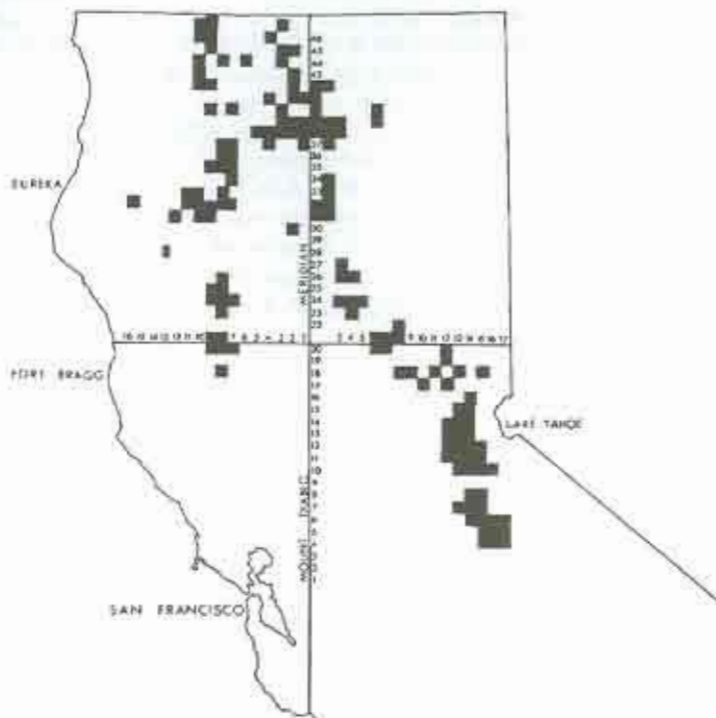


Figure 1. Location of permanent plots by township.

ESTIMATION PROCEDURES AND DATA REQUIREMENTS

STAG is a microcomputer-based program that uses statistical routines to produce a projection set composed of complete individual tree measurements of DBH, H, HCB, species, and tree expansion factor. There are three main data analysis routines in STAG, each with distinct statistical procedures corresponding to the three different classes of data availability: filling in missing data, converting stand table data, and generating stands from summary statistics. In this section we define missing data estimation techniques for both overstory and understory trees. Overstory trees are those trees greater than a defined threshold value of either 5.5 or 11.0 inches in DBH, whereas understory trees are less than or equal to the threshold DBH value. The species discussed in this article use the species codes given in table 1. These species are classified into eight different species groups during the simulation process, also shown in table 1.

TABLE 1. SPECIES CODES AND NAMES

Spp. code	Common name	Spp. abbrev.	Scientific name	Spp. group no.	Spp. group name
01	ponderosa pine	PP	<i>Pinus ponderosa</i> (Laws.)	1	ponderosa pine (PPG)
02	sugar pine	SP	<i>Pinus lambertiana</i> (Dougl.)	2	sugar pine (SPG)
03	incense cedar	IC	<i>Libocedrus decurrens</i> (Torr.)	3	incense cedar (ICG)
04	Douglas-fir	DF	<i>Pseudotsuga menziesii</i> (Mirb.) Franco	4	Douglas-fir (DFG)
05	white fir	WF	<i>Abies concolor</i> (Gordl. and Glend.) Lindl.	5	white fir (WFG)
06	red fir	RF	<i>Abies magnifica</i> (A. Murr.)	6	red fir (RFG)
07	lodgepole pine	LP	<i>Pinus contorta</i> (Dougl.)	1	PPG
08	white pine	WP	<i>Pinus monticola</i> (Dougl.)	2	SPG
09	Jeffrey pine	JP	<i>Pinus jeffreyi</i> (Grev. & Ball.)	1	PPG
10	misc. conifers	CM	n.a.	4	DFG
11	chinquapin	CH	<i>Castanopsis chrysophylla</i> (Dougl.) A. DC.	7	other hardwoods (OHG)*
12	black oak	BO	<i>Quercus kelloggii</i> (Newb.)	8	black oak (BOG)
13	tan oak	TO	<i>Lithocarpus densiflorus</i> (Hook. & Arn.)	7	OHG
14	misc. hardwoods	HM	n.a.	7	OHG

*The OHG equations were derived mainly from CH (11) and TO (13).

FILLING IN MISSING DATA

One of the major and most appropriate uses of STAG is to fill in missing data. The most reliable use of the missing data routines is in combination with statistically valid field data plots. Typically, a statistical sample of trees in a stand type is measured using a series of fixed-area inventory plots of 0.2 or 0.1 acre in size, and trees on those plots are measured for DBH, H, HCB, species, and tree expansion factor. STAG can then be used to fill in occasional missing data values or to provide estimates of variables that may have been subsampled, such as height. It is also possible to fill in missing data values for trees measured on variable-radius (prism) plots. This is not recommended if the data are being prepared for CACTOS, because variable-radius plots typically sample relatively

few trees (4 to 8) at a sample location, compared with fixed-radius plots. It is generally true that DBHs measured on fixed-area plots provide better approximations to diameter distributions than DBHs measured on prism plots, because in the former more trees are sampled. With variable-radius plots, trees are selected with probability proportional to size, and hence larger trees (the trees with greater volume) are more frequently selected. Variable-radius plots are efficient for estimating volume but not for representing the total diameter distribution by species of trees in the stand.

In addition, in prism sampling it is common to measure the DBHs of the sampled trees, but none or few of the sampled trees are actually measured for H or HCB. If missing values are filled in on prism plots, where, say, only DBHs and species have been recorded, the utility of these data for simulation is difficult to assess. Because of these problems, using filled-in data from variable-radius plots for simulation in CACTOS may result in less accurate portrayals of actual stand conditions, and thus yield poorer forecasts than those obtained using fixed-area sample data.

One way to overcome the disadvantages of variable-radius plots for which few tree attributes are measured is to take enough variable-radius plots to ensure that they provide a good approximation of the true diameter distributions for the species present in the stand. The prism plot stand descriptions should be averaged (using the stand description averager (Meerschaert and Wensel (1987)) and then STAG can be used to fill in missing data for the average stand description or the distributional apportionment routines in STAG can be used to convert diameter class data to pseudo-individual tree data. It is also a good idea to supplement the prism plots with small fixed area plots (e.g. $\frac{1}{500}$ th or $\frac{1}{1000}$ th ac) located at the center of each prism plot to better estimate regeneration and small tree frequency.

Estimating Total Height

STAG can be used to fill in tree heights or heights-to-crown base, when one or both are missing, provided that the species, DBH, and expansion factors exist for all trees on the plot. Models 1 and 2 are used to estimate missing heights for overstory (> 5.5 inches DBH) and understory trees (\leq 5.5 inches DBH), respectively. Heights for overstory trees whose diameters exceed 5.5 inches are estimated as a function of DBH, stand basal area, and elevation, as follows:

$$\hat{H}_O = b_0 + b_1 \times \sqrt{\text{DBH}} + b_2 \times \sqrt{\text{BA}_6} + b_3 \times E^2 \quad [1]$$

where \hat{H}_O = the estimated total height (ft) for overstory (o) trees

BA_6 = the stand basal area (ft^2) in trees greater than 5.5 inches in DBH

DBH = tree diameter at breast height (DBH > 5.5 in)

E = stand elevation in feet.

The coefficients b_0 , b_1 , b_2 , and b_3 were estimated for species groups 1 through 8 (see table 1) and an all-species-combined category. Sample sizes for each species group ranged from a low of 340 observations for black oak to over 4,000 observations for ponderosa pine and white fir. All standard errors of prediction

TABLE 2. COEFFICIENTS AND FIT STATISTICS FOR THE TOTAL HEIGHT MODEL 1 FOR OVERSTORY TREES*

Species group and number	$S_{y,x}$	n	b_0	b_1	b_2	b_3
PPG [1]	12.144	4,173	-38.673 (1.234)	27.073 (0.230)	1.809 (0.064)	-7×10^{-7} (3×10^{-8})
SPG [2]	11.215	1,070	-36.456 (2.035)	28.328 (0.353)	0.999 (0.106)	-6×10^{-7} (5×10^{-8})
ICC [3]	9.406	2,260	-28.246 (1.292)	22.713 (0.231)	0.709 (0.065)	-6×10^{-7} (3×10^{-8})
DFG [4]	11.488	2,458	-34.586 (1.482)	27.400 (0.305)	1.446 (0.084)	-6×10^{-7} (4×10^{-8})
WFG [5]	10.700	5,167	-40.147 (1.015)	29.353 (0.186)	0.829 (0.048)	-4×10^{-7} (2×10^{-8})
RFG [6]	11.397	501	-36.656 (3.722)	28.605 (0.558)	1.005 (0.163)	-5×10^{-7} (8×10^{-8})
OHG [7]	13.218	273	-38.731 (5.679)	15.614 (1.270)	2.621 (0.217)	0 (0)
BOG [8]	14.421	340	-2.386 (4.561)	13.237 (0.910)	1.712 (0.260)	-8×10^{-7} (12×10^{-8})
All	13.488	16,242	-35.361 (0.652)	27.607 (0.127)	1.033 (0.034)	-6×10^{-7} (1×10^{-8})

*Standard errors of coefficients shown parenthetically.

were in the range of 9 to 14 feet. Other model forms that included site index were evaluated but did not outperform this model. Coefficient values and fit statistics are presented in table 2.

One model that can be used to predict heights of understory trees is

$$\hat{H}_{uo} = 4.5 + \frac{\hat{H}_{5.5} - 4.5}{5.5} \times \text{DBH} \quad [2]$$

where \hat{H}_{uo} = the estimated total height (ft) for understory trees with an overstory (o) present whose diameter is in the range $0 \leq \text{DBH} \leq 5.5$ inches

$\hat{H}_{5.5}$ = the predicted height (ft) of a 5.5 inch DBH tree from equation 1.

Model 2 simply constrains the predicted height of understory trees to be between 4.5 feet and the height of a 5.5-inch DBH tree as predicted by model 1. This constrained equation can be selected if the user wants to ensure that the understory height predictions smoothly join the overstory equation. This procedure introduces some bias, which can be seen when inspecting the standard error of the regression values for model 2 presented in table 3. The $S_{y,x}$ values for model 2 are considerably larger in 7 out of 8 cases than those of the unconstrained understory height prediction model 3 presented below.

When there is no overstory component, the height of the understory trees can be predicted with equation 3. This equation does not ensure compatibility between the overstory height equation 1 and the unconstrained understory height prediction model 3. Because model 3 is considerably more precise than model 2, we recommend its use whether or not an overstory is present. Model 3 is specified as

$$\hat{H}_u = 4.5 + b_1 \times \text{DBH}^{b_2} \times e^{(b_3 \times E)} \quad [3]$$

where \hat{H}_u = the estimated total height (ft) for an understory tree whose diameter is in the range $0 \leq \text{DBH} \leq 5.5$ inches with no overstory present.

b_1, b_2, b_3 = coefficients estimated for each species group

Coefficients and fit statistics for model 3 are presented in table 3.

TABLE 3. COEFFICIENTS AND FIT STATISTICS FOR THE HEIGHT OF UNDERSTORY TREES MODEL 3 WITHOUT AN OVERSTORY COMPONENT*

Species group and number	$S_{y,x}$ model 3	$S_{y,x}$ model 2	n	b_1	b_2	b_3
PPG [1]	6.109	6.536	1,377	8.621 (0.622)	1.190 (0.035)	-0.00019 (0.00001)
SPG [2]	4.819	5.821	225	3.913 (0.653)	1.287 (0.086)	-0.00008 (0.00003)
ICG [3]	5.436	4.355	996	4.826 (0.379)	1.231 (0.033)	-0.00016 (0.00001)
DFG [4]	5.761	6.466	961	8.431 (0.500)	1.047 (0.030)	-0.00010 (0.00001)
WFG [5]	4.039	5.242	2,471	4.067 (0.191)	1.215 (0.019)	-0.00005 (0.00001)
RFG [6]	4.404	5.299	131	1.590 (0.307)	1.554 (0.127)	0 (0)
OHG [7]	9.591	7.662	76	4.188 (1.604)	1.089 (0.259)	0 (0)
BOG [8]	6.488	7.720	90	6.556 (2.980)	1.187 (0.241)	-0.00019 (0.00005)
All	5.590	6.583	6,327	7.157 (0.224)	1.140 (0.016)	-0.00015 (0.00000)

*Standard errors of coefficients shown parenthetically.

Estimating Height-to-Crown Base

To estimate height-to-crown base (HCB) for overstory trees with DBHs > 5.5 inches, a model form based on the logistic equation was chosen so that HCB would be constrained to be between zero and total height. The form of the model selected was

$$\widehat{\text{HCB}}_O = H \times (1 - e^{-(c_0 + c_1 \times \ln \text{BA}_6 + c_2 \times (\text{DBH}/H)^2)}) \quad [4]$$

where $\widehat{\text{HCB}}_O$ = predicted height (ft) to the base of the crown for overstory (o) trees (> 5.5 in DBH)

H = total height (ft)

DBH = diameter at breast height (nearest 0.1 in)

c_0, c_1, c_2 = coefficients estimated for each species group

BA_6 is as defined in model 1.

We squared the exponential argument to give positivity, which constrains the exponent term to be negative. This constraint ensures that the model always produces a prediction less than the height of the tree. Sample sizes were the same as in estimating the total height model, but standard errors of prediction were slightly less, ranging between 9 and 11 feet. Coefficients and fit statistics are presented in table 4.

TABLE 4. COEFFICIENTS AND FIT STATISTICS FOR THE HEIGHT-TO-CROWN BASE MODEL 4 FOR OVERSTORY TREES*

Species group and number	$S_{y.x}$	n	c_0	c_1	c_2
PPG [1]	10.375	4,173	1.027 (0.035)	-0.112 (0.006)	1.925 (0.060)
SPG [2]	9.454	1,070	1.222 (0.056)	-0.130 (0.009)	1.400 (0.100)
ICG [3]	8.703	2,260	1.119 (0.058)	-0.097 (0.010)	0.974 (0.070)
DFG [4]	11.140	2,458	1.369 (0.051)	-0.162 (0.009)	1.833 (0.095)
WFG [5]	10.856	5,167	1.298 (0.038)	-0.154 (0.006)	1.831 (0.070)
RFG [6]	11.089	501	1.450 (0.113)	-0.160 (0.019)	1.022 (0.175)
OHG [7]	9.188	273	1.727 (0.202)	-0.184 (0.033)	0.535 (0.210)
BOG [8]	10.315	340	1.313 (0.150)	-0.133 (0.028)	0.745 (0.131)
All	10.580	16,242	1.323 (0.019)	-0.146 (0.003)	1.414 (0.030)

*Standard errors of coefficients shown parenthetically.

The prediction equation for height-to-crown base (HCB) of understory trees when an overstory component (o) is present is

$$\widehat{HCB}_{uo} = c_0 + c_1 \times DBH + c_2 \times H + c_3 \times N_6 \quad [5]$$

where \widehat{HCB}_{uo} = predicted height (ft) to the base of the crown for understory trees (≤ 5.5 in DBH) when an overstory (o) is present
 H = total height (ft)
 N_6 = number of trees per acre with DBH > 5.5 inches
 c_0, c_1, c_2, c_3 = coefficients estimated for each species group.

Coefficients and fit statistics for model [5] are presented in Table 5.

For model 5 we observed that the variance increased with increasing predictions of height-to-crown base. We formulated a simple model for this relationship as

TABLE 5. COEFFICIENTS AND FIT STATISTICS FOR THE HEIGHT-TO-CROWN BASE MODEL 5 FOR UNDERSTORY TREES WITH AN OVERSTORY COMPONENT*

Species group and number	S_{yx}	n	c_0	c_1	c_2	c_3	b
PPG [1]	3.393	1,377	2.727 (0.308)	1.737 (0.110)	0.166 (0.016)	-0.0181 (0.0013)	1.3971 (0.0616)
SPG [2]	3.512	224	4.214 (0.871)	1.110 (0.342)	0.252 (0.054)	-0.0192 (0.0028)	1.2756 (0.1129)
ICG [3]	3.697	996	1.764 (0.331)	0.894 (0.134)	0.197 (0.027)	-0.0069 (0.0011)	1.5468 (0.0657)
DFG [4]	4.753	960	1.659 (0.607)	2.567 (0.218)	0.188 (0.027)	-0.0168 (0.0023)	2.0209 (0.0855)
WFG [5]	3.901	2,470	0.866 (0.269)	1.468 (0.119)	0.295 (0.020)	-0.0129 (0.0009)	1.9335 (0.0500)
RFG [6]	3.741	131	3.361 (1.208)	0.437 (0.476)	0.294 (0.083)	-0.0132 (0.0034)	2.0591 (0.2635)
OHG [7]	3.909	75	9.145 (2.005)	-0.599 (0.500)	0.411 (0.008)	-0.0163 (0.0560)	1.0384 (0.1400)
BOG [8]	4.195	90	0 (0)	0 (0)	0.516 (0.044)	-0.0170 (0.0069)	2.2357 (0.3708)
All	3.988	6,323	1.922 (0.173)	1.201 (0.061)	0.302 (0.009)	-0.0159 (0.0006)	1.9789 (0.0347)

*Standard errors of coefficients shown parenthetically.

$$\sigma_{i_{uo}}^2 = b \times \widehat{HCB}_{i_{uo}} \quad [6]$$

where $\widehat{HCB}_{i_{uo}}$ = predicted height (ft) to the base of the crown for the i^{th} understory tree (≤ 5.5 in DBH) when an overstory (o) is present
 $\sigma_{i_{uo}}^2$ = the variance around the regression of the height-to-crown base model 5 for the i^{th} understory tree ($i = 1$ to n)
 b = a coefficient estimated for each species group.

The estimated values of beta (b) in equation 6 are given in table 5.

The procedure for adding stochastic errors is discussed in a later section. Briefly, we predict height-to-crown base for understory trees with DBHs ≤ 5.5 inches using equation 5 or 7. Stochastic errors are then added to the prediction.

When there is no overstory component, we predict the height-to-crown base of understory trees with equation 7 as

$$\widehat{HCB}_u = b_1 \times H^{b_2} + b_3 \times E \times DBH^{b_4} \quad [7]$$

where \widehat{HCB}_u = the predicted height (ft) to the base of the crown for understory trees (≤ 5.5 in DBH) with no overstory present.
 b_1, b_2, b_3, b_4 = coefficients estimated for each species group

Coefficients and fit statistics for model 7 are presented in table 6.

TABLE 6. COEFFICIENTS AND FIT STATISTICS FOR THE HEIGHT-TO-CROWN BASE MODEL 7 FOR UNDERSTORY TREES WITHOUT AN OVERSTORY COMPONENT*

Species group and number	$S_{y,x}$	n	b_1	b_2	b_3	b_4
PPG [1]	3.507	1,377	0.964 (0.321)	0.534 (0.084)	0.00017 (0.00006)	1.372 (0.181)
SPG [2]	4.046	225	1.400 (0.320)	0.664 (0.074)	0 (0)	0 (0)
ICG [3]	3.080	996	0.203 (0.019)	0 (0)	0.00033 (0.00004)	0.699 (0.087)
DFG [4]	4.914	961	0.254 (0.017)	0 (0)	0.00034 (0.00006)	1.219 (0.095)
WFG [5]	4.189	2,471	0.615 (0.047)	0.920 (0.025)	0 (0)	0 (0)
RFG [6]	3.971	131	1.242 (0.418)	0.635 (0.112)	0 (0)	0 (0)
OHG [7]	4.209	76	1.358 (0.309)	0.752 (0.068)	0 (0)	0 (0)
BOG [8]	4.222	90	0.965 (0.381)	0.742 (0.124)	0 (0)	0 (0)
All	4.214	6,327	0.610 (0.080)	0.830 (0.034)	0.00006 (0.00002)	1.608 (0.206)

*Standard errors of coefficients shown parenthetically.

It can be seen by comparing $S_{y,x}$ in tables 5 and 6 that when we know the sizes and numbers of trees in the overstory the precision of the height-to-crown base models increases by about five percent.

With these equations it is possible to fill in or estimate missing values of height and height-to-crown base for individual trees. The only exogenous variable that needs to be supplied for each stand is elevation. Basal area (BA_0) can easily be computed directly by summing the per acre individual tree basal areas obtained from the individual tree DBHs and expansion factors contained in the projection set file (stand description file). Number of trees (N_0) can easily be calculated from the expansion factors associated with individual trees in the projection set file.

Converting Merchantable Height to Total Height

The four different types of tree height measurements allowed in STAG include: (1) total heights, (2) heights to a merchantable top (≤ 6.5 inches diameter inside bark), (3) heights measured to whole (16.5 ft) logs, or (4) heights measured to half logs (8.25 ft). Within a STAG projection set file, which is composed of individual tree measurements, all heights must have the same measurement standard. CACTOS requires total heights for individual trees, but STAG can manipulate merchantable height to obtain an estimate of total height. STAG uses a taper equation to solve for total height for the six major conifer species (species group numbers 1–6 in table 1) whenever height to a merchantable top or number of 16.5-foot logs is supplied.⁴

⁴The height conversion process is not intended to encourage the measurement of other than total heights. Rather, it is intended to allow the use of older inventory data.

We derive total height (H) from merchantable height (MH) with a sigmoid taper equation (Biging 1984). The total height estimate obtained from inverting the taper equation is

$$\hat{H} = \frac{MH \times (\lambda)^3}{(1 - \exp[(d/DBH - b_1) / b_2])^{b_3}} \quad [8]$$

where \hat{H} = the predicted total tree height (ft) estimated from merchantable height
 λ = $1 - \exp(-b_1/b_2)$
 DBH = the diameter at breast height (in)
 d = the merchantable top diameter (≤ 6.5 in)
 MH = the height to the merchantable top diameter (ft)
 $\exp(x)$ = 2.71828...raised to a power of x
 b_1, b_2 = species-specific coefficients given in table 7.

Coefficients and fit statistics for model 8 taken from Biging (1984) are presented in table 7.

TABLE 7. COEFFICIENT ESTIMATES BY SPECIES
FOR EQUATION 8 FROM BIGING (1984)

Species and spp. codes	n	b_1	b_2
PP [1]	2,014	1.019589	0.335666
SP [2]	692	1.069320	0.415632
IC [3]	541	1.071343	0.472157
DF [4]	1,588	1.029288	0.334012
WF [5]	2,645	1.092615	0.365295
RF [6]	312	1.075880	0.353784

If the heights of trees are entered as the number of logs, the program first converts them to heights to the given merchantable top using equation 9, and then uses equation 8 to predict total heights. The equation to estimate height to the merchantable top (MH) when only the number of logs is known is given by

$$\widehat{MH} = \left[SH + \frac{LL}{2} \right] + NLOGS \times LL \quad [9]$$

where \widehat{MH} = the estimated height to the merchantable top diameter (ft)
 SH = stump height of the tree = 1.5 feet
 LL = log length in feet (16.5 or 8.25 ft)
 NLOGS = the number of logs of length LL for a tree.

Stochastic Errors

The user can either make a deterministic or a stochastic prediction of missing values when filling in missing data (height or height-to-crown base) or generating stands from summary statistics. Choosing stochastic errors means that a random

value will be added to the prediction to reflect that an individual tree's dimensions cannot be predicted with certainty. Thus, a random value will be added or subtracted from the prediction. For the overstory tree height model 1, the understory tree height model 2, the understory height model with no overstory component (model 3), the overstory height-to-crown base model 4, and the understory height-to-crown base model with no overstory component (model 7), a random value is drawn from a normal distribution with mean zero and variance equal to the estimated variance around the regression (S^2_{YX}). There are two special cases. Even though equation 2 is a constrained model for which no statistical model was fit, we calculated the estimated variance (S^2_{YX}) from model 2 to generate stochastic errors for this understory height equation. See table 3 for estimated (S^2_{YX}) values for model 2. In the case of the understory height-to-crown base model 5, the distributional mean is zero for random errors, but the variance is proportional to the predicted height-to-crown base (see equation 6). If random errors are not requested, then the missing value is set equal to the model prediction (the deterministic prediction). If random errors are not added, all predicted heights and height-to-crown base values are identical for a given diameter of a particular species, given that basal area and elevation are the same.

Parameter Updating

If the user wants to incorporate knowledge of a local sample into the height model coefficients, a Bayesian update of the first two parameters of height model 1 is possible. Alternatively, an ad hoc weighting scheme patterned after the linear composite estimators (Burk, Hansen, and Ek 1982) can be chosen. In both cases, only the first two parameters (b_0 and b_1) are allowed to be updated, because the effects of elevation (E) and density (BA_0) cannot be adequately described with a local sample.

The ad hoc approach adjusts the amount of change to the model parameters by a constant ratio (k) between 0 and 1. A weight of $k = 0$ causes the update routine to abort (no update), while a weight of $k = 1$ places all the emphasis on the local sample to determine the coefficient values to be used for the height prediction equations. This ad hoc weighting process is given as

$$\begin{aligned}\hat{\beta} &= K \times I \times \beta_L + (1-K) \times \beta_D \\ \hat{\beta} &= [b_0^u, b_1^u] \text{ where } b_0^u \text{ and } b_1^u \text{ are the updated parameter estimates}\end{aligned}\quad [10]$$

where

- K = the ad hoc weight matrix (2×2) with diagonal elements (k) $0 \leq k \leq 1$
- I = an identity matrix (2×2)
- β_D = the database matrix estimate of the parameters (2×1)
- β_L = the matrix estimate of the parameters based on the local sample (2×1).

We modified the true Bayesian method because in prior work (Van Deusen 1984) we found that it worked poorly. With relatively small sample sizes the Bayesian update could result in large covariance terms in the local covariance

matrix, which could cause the updated parameter estimates to behave poorly. For example, the local parameter estimates could indicate that both the database slope and intercept coefficients should be increased over their database counterparts. A large negative covariance term in the local covariance matrix could force these two coefficients to move in opposite directions, regardless of the fact that both local parameter estimates were larger than the database estimates. Because of this, we modified the Bayesian approach and have termed it a pseudo-Bayesian approach. The main difference between a Bayesian and a pseudo-Bayesian approach is that for the latter we utilize only the variance terms in the variance-covariance matrices of the local and database samples to avoid problems associated with the covariance terms.

The pseudo-Bayesian approach is more conservative than the ad hoc procedure. If the local sample is small, then the updated coefficients for the height prediction equation are quite close to the database values. If, however, there is a large local sample, then the pseudo-Bayesian estimates are a compromise between the database values and those determined from the local sample. The pseudo-Bayesian update is given by

$$\hat{\beta} = W \times \beta_L + (I - W) \times \beta_D \quad [11]$$

where $\hat{\beta} = [b_{00}^H, b_{10}^H]$, where b_{00}^H and b_{10}^H are the updated parameter estimates

I = an identity matrix (2×2)

β_D = the database estimate of the parameters (2×1)

β_L = the matrix estimate of the parameters based on the local sample (2×1)

W = the weighting matrix (2×2) = $(V_D^{-1} + V_L^{-1})^{-1} V_L^{-1}$

V_D^{-1} = the diagonal elements of the inverse of the variance matrix for the database parameters

V_L^{-1} = the diagonal elements of the inverse of the variance matrix for the parameters based on the local sample.

Van Deusen (1984) found that if the local estimate is of sufficient size it is often the best, but when uncertainty exists the ad hoc or pseudo-Bayesian methods are reliable, with the pseudo-Bayesian being conservative and of low risk.

GENERATING STANDS FROM SUMMARY STATISTICS

In cases where no individual tree measurements are available or when only summary statistics are recorded by species, it is possible to generate a facsimile description of a stand. This facsimile description still meets the qualifications of a projection set because it is composed of complete individual tree records calculated using summary statistics. With knowledge of the summary statistics it is possible to generate a diameter distribution, as developed in a later section. Individual tree diameters can be sampled from this distribution. Tree height and height-to-crown base values are estimated from equations 1 through 7 to complete the facsimile projection set.

The goal of this methodology is to produce a facsimile projection set of

complete individual tree records that is plausible given the specified summary statistics. Users of this technique should be aware that this may produce highly variable results, since there is often a wide range of stand compositions that will have similar summary statistics. We have done limited testing of the stand generation procedures using permanent plot data for mixed-species, multiple-aged coniferous stands and have found, in these test cases, that the stand generation produced "reasonable" facsimiles of stands. Because we tested only a small subset of the possible types of stands that could be generated, we cannot say that this technique generally can be applied with good results. Therefore, the stand generation technique should be used with great caution. We recommend that this technique be used only as a last resort, not as a matter of course. There is no replacement for real field data.

Generation of Overstory Trees

Overstory trees are defined as those greater than a specified threshold value (usually 5.5 or 11.0 in DBH), and understory trees are those at or below the threshold value. We have developed separate approaches for generating overstory and understory trees to achieve better accuracy in predicting missing data values.

The joint distribution of species, diameter at breast height, total height, and height-to-crown base is formulated as a product of probability density functions (Van Deusen 1984; Biging and Wensel 1987). This joint probability distribution for overstory trees can be represented as a mixture of distributions:

$$p(\text{DBH}, H, \text{HCB}) = \sum_{\text{species} = 1}^S p(\text{species}) \times p(\text{DBH} | \text{species}) \times p(H | \text{species, DBH}) \times p(\text{HCB} | \text{species, DBH, } H) \quad [12]$$

where S = the number of species present in the stand.

The joint probability distribution of diameter at breast height, total height, and height-to-crown base ($p(\text{DBH}, H, \text{HCB})$) is factored as a product of three conditional distributions. The first term on the right-hand side is $p(\text{species})$, which is the fraction of each species in the stand. This is easily specified by supplying the number of trees per acre by species in the hypothetical stand.

The three conditional distributions are for diameter at breast height, total height, and height-to-crown base. The first of these conditional distributions is that of the diameter of a given species ($p(\text{DBH} | \text{species})$). The conditional diameter distribution can be generated from either a two-parameter truncated Weibull or a negative exponential distribution by relating the summary statistics to the parameters of these distributions. The first two moments of the Weibull distribution correspond to the average diameter of the species and the squared quadratic mean diameter of the species, which can be derived from basal area and number of trees for each species (see equation 16).

We found that the first moment (average diameter for a given species) could be accurately predicted as a function of elevation, the quadratic mean stand

diameter, and numbers of trees in the species. This is discussed more fully in a following section (see equation 15). The user can generate a diameter distribution for each species having knowledge of only the number of trees and basal area in each species. Individual tree DBHs are then randomly generated using an inverse transformation method for either the two-parameter Weibull or the negative exponential.

To randomly sample from this distribution we will consider it a probability density function, compute its associated cumulative distribution function, and finally compute the inverse cumulative distribution function from which we may generate DBHs. Since the cumulative distribution function produces a probability, and by definition probabilities are bounded between 0 and 1, we may use uniformly distributed random deviates bounded between 0 and 1 to generate values for input into the inverse cumulative distribution function.

A second distribution, the negative exponential, was provided for the infrequent case in which a balanced, uneven-aged condition exists within a stand. The details for the procedure of fitting the distribution and a list of the necessary stand summary statistics are provided in a later section.

With either the Weibull or negative exponential distribution, only unimodal distributions can be generated for a given species. In most cases there are too few trees of a given species to develop more complex distributional models. However, because we allow each species to have its own diameter distribution, it is possible to build multimodal distributions for a stand.

Once the diameters are specified with the diameter distribution, the overstory height and height-to-crown base values are predicted with equations 1 and 4. The understory height and height-to-crown base are predicted with either equations 2 and 5, or 3 and 5, or 7. These equations are associated with the conditional probability distributions of $p(H|species, DBH)$ and $p(HCB|species, DBH, H)$, respectively. Elevation also needs to be supplied, since it is an independent variable in height prediction equation 1 and in understory height-to-crown base equation 7. Random stochastic errors distributed as $N(0, S_{\epsilon}^2)$ for equations 1, 2, 3, and 4 or as $N(0, b \times HCB_{base})$ for equations 5 and 7 are added to the predictions if the random error feature has been selected.

There are alternatives to the factorization approach utilized in this study. For example, the joint distribution of diameter, height, and height-to-crown base could have been modeled as a trivariate distribution. We did not investigate this approach because we had relatively few measured trees on each of the re-measured permanent plots (usually less than 20). With a factorization approach there is the additional advantage that any number of species can be modeled.

Weibull Distribution

The Weibull distribution has been widely used in forestry applications for describing the diameter distributions of stands. This use stems both from the Weibull's shape and the ease of estimating parameters. We chose a truncated Weibull because we improved our ability to model diameter distributions by breaking the stand into overstory ($DBH > 5.5$ in) and understory components ($0 \leq DBH \leq 5.5$ in). The three-parameter Weibull may be reduced to the two-parameter Weibull since the location parameter, typically called a , is zero. The two-parameter truncated Weibull density function is given as (Van Deusen 1984)

$$f(x) = \left[\frac{c}{b} \right] \times \left[\frac{x}{b} \right]^{c-1} \times e^{-(x/b)^c} \quad [13]$$

where: $f(x)$ = frequency of trees in diameter class x
 x = midpoint DBH of diameter class; $x \geq T$
 T = truncation DBH (5.5 in)
 b and c = parameters > 0 .

Deriving the b and c coefficients

To specify a particular distribution from this Weibull family, we need to define the b and c parameters. The moment equation for the two-parameter truncated Weibull is given as:

$$Ex^r = b^r \times c^{(T/b)^c} \times \left[\Gamma(r/c+1) - c \times b^{-(r/c)} \times \int_0^T x^{r+c-1} \times e^{-(x/b)^c} \times dx \right] \quad [14]$$

where: Ex^r = the expectation of the r^{th} moment of x or DBH
 $\Gamma(r/c+1)$ = the gamma function of $(r/c + 1)$.

The first and second moments are used to simultaneously solve for the b and c parameters. It is known from Cauchy's inequality that the arithmetic mean must be less than or equal to the quadratic mean stand diameter, which is the square root of equation 16. The arithmetic mean stand diameter is predicted as a fraction of the quadratic mean stand diameter, where the fraction is constrained to be less than 1 through use of the logistic function. The arithmetic mean stand diameter is predicted as follows:

$$D^{(1)} = \overline{\text{DBH}} = \left[\beta_0 + \frac{(1 - \beta_0)}{(1 + e^{(-\beta_1 - \beta_2 \times E - \beta_3 \times \ln(\overline{D}_q) - \beta_4 \times \text{SN}_6 - \beta_5 \times \overline{D}_q^{-1}))})} \right] \times \overline{D}_q \quad [15]$$

where: $D^{(1)}$ = the estimated $\overline{\text{DBH}}$ = the first moment or mean stand diameter for a given species
 \overline{D}_q = quadratic mean diameter of trees for a given species > 5.5 inches DBH
 SN_6 = number of trees for a given species > 5.5 inches DBH
 E = elevation (ft)
 β_0, \dots, β_5 = the coefficients estimated from regression (see table 8).
 \ln = logarithm base e .

TABLE 8. COEFFICIENTS AND FIT STATISTICS FOR MEAN STAND DIAMETER MODEL 15 FOR ALL SPECIES(1-8) COMBINED*

Number of plots	MSE	β_0	β_1	β_2	β_3	β_4	β_5
2,078	0.363	0.75637 (0.24924)	-12.12687 (6.08516)	-0.00018 (0.00008)	3.62041 (1.31558)	6.15495 (3.09834)	56.31421 (20.51430)

*Standard errors of coefficients shown parenthetically.

The second moment is the squared quadratic mean stand diameter, which is given by definition as

$$D^{(2)} = \bar{D}_q^2 = \frac{1}{n} \times \sum DBH_i^2 = \frac{SBA_6}{K' \times SN_6} \quad [16]$$

where $D^{(2)}$ = the estimated second moment or squared quadratic mean stand diameter (\bar{D}_q^2) of a given species
 n = the number of trees on a plot
 K' = 0.005454, which is a conversion factor for diameter in square inches to basal area in square feet
 SBA_6 = basal area of trees for a given species > 5.5 inches DBH
 SN_6 = number of trees for a given species > 5.5 inches DBH.

The absolute difference between the two moments given above and the predicted moments given an estimated b and c provides an overall error figure. This figure is required to be less than 5 percent ($E\%$) of the respective moments. The square-root transform is used on the second moment, so that it can be put into the same metric as the first moment. The error formula is

$$E = \frac{E\% \times D^{(1)} + E\% \times \sqrt{D^{(2)}}}{2} \quad [17]$$

where E = the average error between predicted and observed moments (inches)
 $E\%$ = the percent error allowed in estimating the moment
 $D^{(1)}, D^{(2)}$ = as defined above.

Using ponderosa pine as an example, if the species basal area were given as 150 ft^2 and the species number of trees per acre given as 300, then the maximum allowable error in finding the Weibull parameters would be

$$E = \frac{(0.05 \times 9.01 + 0.05 \times \sqrt{91.68})}{2} = 0.46 \text{ inches}$$

where 9.01 is $D^{(1)}$ and 91.68 is $D^{(2)}$. When estimates of b and c are obtained, then predicted values of $\hat{D}^{(1)}$ and $\hat{D}^{(2)}$ are obtained and summed as

$$\hat{E} = \frac{9.01 - \hat{D}^{(1)} + \sqrt{91.68 - \hat{D}^{(2)}}}{2}$$

where $\hat{D}^{(1)}$ and $\hat{D}^{(2)}$ are the predicted first and second moment, respectively. If \hat{E} is less than E , then the estimates of b and c are close enough and the procedure stops; otherwise, new estimates of the parameters are calculated and the process is repeated.

To begin the algorithm for determining the coefficients, two predictive equations are used to provide starting values for the parameters. These equations were fit using multiple linear regression for converged values:

$$\hat{b} = -2.2675 - 0.5433 \times D^{(1)} + 0.01304 \times D^{(2)} + 0.0002460 \times D^{(1)} \times D^{(2)} - 0.01175 \times (D^{(1)})^2 - 0.000001542 (D^{(2)})^2$$

$$\hat{c} = \exp(-0.7970 - 0.6513 \times \hat{b} + 0.2501 \times \hat{b}^2 - 0.6576 \times D^{(1)} \times \hat{b} + 0.001782 \times D^{(2)} \times \hat{b} + 1.0459 \times D^{(1)} - 0.006255 \times D^{(2)} - 0.00175 \times D^{(1)} \times D^{(2)} + 0.4085 \times (D^{(1)})^2)$$

Next, Powell's method is used to minimize \hat{E} . If Powell's method fails to converge within 12 iterations, then a grid search is used to minimize \hat{E} . Powell's method is much faster than a grid search and is quick to converge for most stands with mean stand diameters of about 6 inches DBH and above.

The grid search algorithm begins by searching for the minimum error over a coarse grid (increment of 0.5) with respect to \hat{b} and \hat{c} . This coarse grid search is accelerated by retrieving the $\hat{D}^{(1)}$ and $\hat{D}^{(2)}$ estimates from two binary files. Next, the range of the parameters is reduced to be around the minimum found in the coarse grid search, the increments for \hat{b} and \hat{c} are reduced to a third of their previous value, and a finer search is performed. Up to 10 iterations of increasingly finer grid searches are performed. As with Powell's method, the convergence criteria is that \hat{E} be less than E . Once the grid search converges, a fine tuning is performed where \hat{b} and \hat{c} are adjusted slightly, so that the relative error in estimating the first moment is approximately the same as that of the second moment.

Estimating DBHs

The derivation for the inverse cumulative distribution function of the truncated two-parameter Weibull is as follows. The probability density function is integrated from the lower truncation point (T) to the diameter of interest (x):

$$F_T(x) = \int_T^x \left[\frac{c}{b} \right] \times \left[\frac{t}{b} \right]^{c-1} \times e^{[(T/b)^c - (t/b)^c]} dt = [1 - e^{-(x/b)^c} \times e^{-(T/b)^c}] \quad [18]$$

where $F_T(x)$ = the cumulative proportion of trees between the lower truncation point (T) and the specified upper diameter (x) with $0 \leq F_T(x) \leq 1$

b and c = estimated constants.

This provides the cumulative proportion of trees up to the diameter of interest. Generating a uniform random number between 0 and 1 gives us a value for $F_T(x)$; we may then solve for the DBH by inverting equation 18 as follows:

$$DBH = b \times [(T/b)^c - \log_e (1 - F_T(x))]^{1/c} \quad [19]$$

By generating a uniform random number ($F_T(x)$), we can use the inverse transform of the cumulative distribution function to estimate diameters at breast height using equation 19.

Negative Exponential Distribution

The diameter distributions of balanced uneven-aged stands (Meyer 1952) are often characterized as being distributed according to the negative exponential distribution. A typical method for applying the distribution to a stand is with the diminution quotient or Q value (Husch, Miller, and Beers 1982; Davis and Johnson 1987). To obtain the number of trees in the next to the largest diameter class, we would simply multiply Q by the number of trees in the largest diameter class. Thus, for the next smallest diameter class, we would multiply Q times the number of trees in the next largest diameter class or Q^2 times the number of trees in the largest diameter class. To compute the number of trees in each diameter class, we need to specify Q , a range of tree diameters, their diameter class (≥ 1 in), and the number of trees in the largest diameter class. Unlike with the truncated Weibull distribution, we use the negative exponential distribution to simultaneously generate both overstory and understory trees.

The negative exponential function that is used to describe the distribution of numbers of trees by diameter class is given as follows:

$$\widehat{SN}_i = k \times e^{-a \times DC_i} \quad [20]$$

where \widehat{SN}_i = the estimated number of trees of a given species in i^{th} diameter class ($i = 1, \dots, n$)
 DC_i = the i^{th} diameter class
 a and k = coefficients.

To specify the distribution we need to define k and a , which can be done by using Q , a value that may be more meaningful to a manager than the a and k coefficients of the negative exponential function.

Deriving the a and k coefficients

The coefficient a is derived from the definition of Q :

$$Q = \frac{\widehat{SN}_{n-1}}{\widehat{SN}_n} = \frac{k \times e^{-a \times DC_{n-1}}}{k \times e^{-a \times DC_n}} \quad [21]$$

where Q = the estimated diminution quotient
 \widehat{SN}_{n-1} = the estimated number of trees for a species in the next to the largest diameter class
 \widehat{SN}_n = the estimated number of trees for a species in the largest diameter class
 DC_{n-1} = the next to the largest diameter class of a given species
 DC_n = the largest diameter class of a given species

C = the size in inches (width) of the diameter class of a given species that equals $DC_n - DC_{n-1}$.

Solving for a we get

$$\hat{a} = \frac{\log \hat{Q}}{C} \quad [22]$$

Since we know the number of trees in the largest diameter class (SN_n) and the a parameter, we may solve for k using the negative exponential equation

$$\hat{k} = \frac{\widehat{SN}_n}{e^{-a \times DC_n}} \quad [23]$$

When Q and SN_n are known, we can estimate the a and k parameters needed for the negative exponential distribution in equation 20.

Calculations when Q or SN_n is unknown

If either Q or SN_n is unknown, then the basal area for the species on the plot (SBA) is used to compute the missing variable.

If Q and species basal area (SBA) are known but SN_n is unknown, we iteratively solve for SN_n using equation 24. In equation 24 species basal area is formulated as the sum of the number of trees in a diameter class multiplied by the square of the diameter class.

$$SBA = K' \times \sum_{i=1}^n \widehat{SN}_i \times DBH_i^2 \quad [24]$$

where SBA = total basal area for the species on the plot in square feet
 \widehat{SN}_i = the estimated number of trees of the species in the i^{th} diameter class ($i=1, \dots, n$)
 DBH_i = the midpoint DBH of the i^{th} diameter class in inches for the species
 K' = 0.005454, which is a conversion factor for diameter in square inches to basal area in square feet.

To estimate SN_n from Q and SBA, we initially give SN_n a starting value of 1. Then if the estimated SBA is less than the specified SBA, SN_n is increased by 0.001, or vice versa. Using the new estimate of SN_n , the procedure is repeated until the difference between the specified and estimated SBA is less than 1 square foot.

If Q is unknown, then SBA and SN_n must be given. Q is then computed using an iterative process where Q is initially set to 1.1. The number of trees in each diameter class i of a given species is estimated by

$$SN_i = SN_n \times Q^{n-i} \quad [25]$$

where i = is an index for the i^{th} diameter class (1,..., n) ordered from smallest to largest, respectively
 n = the number of diameter classes.

The SBA is then estimated with equation 24 and compared with the specified basal area, and the estimate of Q is incremented identically as the estimated SN_n is incremented above. The same threshold of 1 square foot of basal area difference between the specified and estimated SBA is used as a stopping criterion. Q can then be estimated from equation 21.

Now that all of the necessary information is complete, the coefficients for the negative exponential distribution may be easily computed. The diameters are simulated and written to the projection set file, each with an expansion factor of 1. The total tree height and height-to-crown base are also estimated, given the simulated diameter at breast height using equations 1 to 7. The total number of trees for the species on the plot is rounded to the nearest integer so that all the trees in the completed projection set file will have an expansion factor of 1.

Estimating DBHs

The derivation for the inverse cumulative distribution function of the negative exponential is as follows. The probability density function is integrated over the range of diameters up to the diameter of interest:

$$\begin{aligned} F(\text{DBH}) &= \int_m^{\text{DBH}} \frac{\hat{k}}{C} \times e^{-\hat{a} \times \text{DC}} \times d\text{DC} = \frac{\hat{k}}{C} \times \left[\frac{e^{-\hat{a} \times \text{DC}}}{-\hat{a}} \right]_m^{\text{DBH}} \\ &= - \left[\frac{\hat{k}}{C \times \hat{a}} \right] \times [e^{-\hat{a} \times \text{DBH}} - e^{-\hat{a} \times m}] \end{aligned} \quad [26]$$

where $F(\text{DBH})$ = the cumulative number of trees between the minimum diameter (m) and the specified upper diameter of interest (DBH)

\hat{k} and \hat{a} = estimated constants

m = $\text{DC}_{\min} - \frac{C}{2}$, with C = the class width in inches

DC_{\min} = minimum diameter class in inches.

This provides the cumulative number of trees up to the diameter of interest. Generating a uniform random number between 0 and 1 and multiplying it by the total number of trees gives us a value for $F(\text{DBH})$. We may then solve for the DBH by inverting the above equation, as is seen in equation 27

$$DBH = \frac{-\ln(-F(DBH) \times (\frac{C \times \hat{a}}{\hat{k}}) + e^{-\hat{a} \times m})}{\hat{a}} \quad [27]$$

Thus, by generating a uniform random number ($F(DBH)$) we can use the inverse transform of the cumulative distribution function to estimate diameters at breast height.

Generation of Understory Trees

As an adjunct to the stand generation techniques (overstory generation), we have developed the capability to generate understory trees. The understory trees can be between 1.0 and 11.0 inches at DBH. The overstory of trees (measured or generated) can be used to predict the b and c parameters of the Weibull needed to generate understory trees. Understory tree height and height-to-crown base values are estimated from equations 2, 3, 5, and 7 to complete the understory facsimile projection set. Because stands of trees are often simulated for over 30 years with CACTOS, it is essential to be able to generate an understory component that matures with relatively long simulations. One reason we separated the overstory and understory components is that there is much greater variability (plot to plot or stand to stand) in the number of understory trees than in the number of overstory trees. Hence, the understory generator is inherently more imprecise.

A two-parameter Weibull distribution was fit to the understory component (1.0 in \leq DBH \leq 11.0 in) for each of the 308 permanent plots for which there were at least six understory trees present on the plot with a plot average diameter exceeding 5.5 inches. Six trees was chosen as the minimum number needed for estimating the two parameters of the Weibull distribution, although most plots had many more than six trees. An 11.0-inch DBH was chosen as an upper value for the distribution rather than a 5.5-inch DBH value to allow for a more regular distributional form and to increase the number of trees available for modeling the understory diameter distribution. Even though an upper DBH value of 11.0 inches was chosen, the understory generation can be specified for any range within these limits. Summary statistics for the understory component of the 308 permanent plots used to model the Weibull diameter distribution are presented in Appendix B.

We found that the coefficients of the Weibull for the understory could be predicted as the following functions of overstory parameters:

$$\hat{b} = b_0 + \frac{b_1}{N_6} + \frac{b_2}{DBH_{min}} + b_3 \times CV_6 + \frac{b_4 \times CV_6}{DBH_{min}} \quad [28]$$

$$\hat{c} = c_1 \exp(c_2 \times \hat{b} + c_3 \times X^c + c_5 \times Y + c_6 \times Z) \quad [29]$$

where \hat{b} = predicted value for b (scale) parameter of two-parameter, left-truncated Weibull

- \hat{c} = predicted value for c (shape) parameter of two-parameter, left-truncated Weibull
- N_6 = the number of trees per acre greater than 5.5 inches in DBH
- DBH_{min} = the minimum diameter measured on a specific plot, usually 1.0 or 2.0 inches
- CV_6 = the coefficient of variation of DBH for trees greater than 5.5 inches DBH
- X = $0.75 + BA_6/87.945 - SDI_6/131.0$
- BA_6 = stand basal area in trees greater than 5.5 inches DBH
- SDI_6 = stand density index considering only trees greater than 5.5 inches DBH (cf. Reineke 1933; Avery and Burkhart 1983)
- Y = $0.035 + 1/(b - SDI_6)$
- Z = $410.0 + 1/[\ln(\hat{b}) - \ln(BA_6)]$
- b_0, \dots, b_4 = b coefficients estimated for all species combined
- c_1, \dots, c_6 = c coefficients estimated for all species combined.

Coefficient values and fit statistics appear in table 9. Due to the great inherent variability of the understory component, these predictive equations explain a small but significant portion of the total variability. While the predicted parameters resulting from using these equations will not be very precise, they are still preferred over using a simple average value. In general, predicting the parameters of a Weibull distribution from stand characteristics, even in situations in which there is not a great deal of variation, has proven difficult, and R^2 values are typically less than 0.10 (Knobel and Burkhart 1991).

TABLE 9. COEFFICIENTS AND FIT STATISTICS
FOR THE UNDERSTORY WEIBULL PARAMETERS OF MODELS 28 AND 29
FOR ALL SPECIES COMBINED ESTIMATED USING 308 PLOTS*

Model	MSE	b_0	b_1	b_2	b_3	b_4	
[28]	3.718	15.4269 (1.9121)	-66.160 (18.702)	-15.0426 (3.5167)	-0.13859 (0.03925)	0.21027 (0.07143)	
Model	MSE	c_1	c_2	c_3	c_4	c_5	c_6
[29]	3.691	0.57718 (0.25234)†	0.31116 (0.04067)	0.51325 (0.15300)	2.9000 (0)‡	-18.8181 (8.6527)	4.0501×10^{-5} (1.9817×10^{-6})

*Standard errors for coefficients shown parenthetically.

†Standard error for $\ln(c_1)$

‡The c_4 parameter was fixed at a value of 2.90 and hence has no standard error.

Specification of total numbers of understory trees

Even though the b and c parameters of the Weibull distribution can be directly predicted via equations 28 and 29, this distribution gives only the relative frequency of tree sizes. Therefore, the total number of understory trees must be specified before the understory can be generated. The total number of understory trees can also be predicted from overstory parameters. Predicting the total number of understory trees from overstory parameters is analogous to predicting the number of ingrowth trees—numbers of trees that will reach some

minimum surveyed size in a specified time—with some obvious differences. They are similar in that both involve only use of overstory conditions to estimate the condition of the understory. The prediction of ingrowth numbers is arguably more well defined than the prediction of total understory numbers, in the sense that one particular size class is under scrutiny, while predicting understory numbers may involve a broad spectrum of size classes. On the other hand, estimates of ingrowth are further complicated by an implied growth rate of trees whose exact sizes are unknown, while estimates of the total number of understory trees represent a static depiction of the stand at one instant in time. Both estimation problems are complicated by the fact that stands currently with similar overstory conditions may have had dissimilar histories, which may result in dissimilar understory conditions.

Models frequently used to predict ingrowth have been reviewed by Shifley (1990). Typically, variables important to the prediction of ingrowth involve stand density measures such as basal area per acre, number of trees per acre, percent stocking, and sum of diameters per acre. These variables also affect growth rates of individual trees, so their superiority in predicting ingrowth is somewhat to be expected. One might also expect that additional variables may be required to predict total number of understory trees due to the previously noted differences between these two estimation problems.

A useful approach to modeling the number of understory trees was found by viewing the problem as the specification of total stand structure based on what was found in the overstory portion only. This led to the investigation of several stand structure variables. Shifley and Lentz (1985) pointed out that the ratio of the mean DBH to the standard deviation of DBH was a valuable index to the *c*, or shape, parameter in the Weibull distribution. Miller and Weiner (1989) and Knox, Pect, and Christensen (1989) found that the inverse of Shifley's index, commonly known as the coefficient of variation, was useful in describing size inequality, or the degree of size hierarchy development in populations of forest trees. We found that the ratio of variance of DBH to the mean DBH was a useful predictor in our models for estimating total number of understory trees.

A model for predicting the number of understory trees was patterned after the ingrowth models of Ek (1974) and Hyink and Moser (1983). The same model form is used for predicting the number of trees between 1.5 and 5.5 inches DBH (N_{1-6}) as for the number of trees between 5.6 inches and 10.5 inches DBH (N_{6-11}). The predictions for understory tree numbers are given by

$$\hat{N}_{1-6} = \exp \{ b_0 + b_1 \times \text{DSUM}_6^{b_2} \times N_6^{b_3} + b_3 \times (R_6 + 1.5) \times N_6^{b_4} \} \quad [30]$$

$$\hat{N}_{6-11} = \exp \{ c_0 + c_1 \times \text{DSUM}_{11}^{c_2} \times N_{11}^{c_3} + c_3 \times (R_{11} + 1.5) \times N_{11}^{c_4} \} \quad [31]$$

- where \hat{N}_{1-6} = the predicted number of trees per acre with 1.5 in \leq DBH \leq 5.5 in
 \hat{N}_{6-11} = the predicted number of trees per acre with 5.5 in $<$ DBH \leq 10.5 in
 N_6 = the number of trees per acre whose DBH $>$ 5.5 inches
 N_{11} = the number of trees per acre whose DBH $>$ 10.5 inches
 R_6 = the ratio of variance of DBH to mean DBH for trees $>$ 5.5 inches

- R_{11} = the ratio of variance of DBH to mean DBH for trees > 10.5 inches
 $DSUM_6$ = the sum of the diameters for trees whose DBH > 5.5 inches
 $DSUM_{11}$ = the sum of the diameters for trees whose DBH > 10.5 inches
 b_0, \dots, b_4 = b coefficients estimated for each forest type (see table 10)
 c_0, \dots, c_4 = c coefficients estimated for each forest type (see table 11).

We found that predictions could be improved through stratification by forest type. An analysis was performed to see if any of the classes could be combined, but we found that a statistically significant improvement was made by using separate coefficients for each major forest type for predicting both N_{1-6} and N_{6-11} . The coefficients were therefore estimated by timber type and are given in tables 10 and 11.

TABLE 10. COEFFICIENTS FOR PREDICTED NUMBER OF UNDERSTORY TREES N_{1-6} * (Model 30)

Timber type	$S_{y.x}$	Number of plots	b_0	b_1	b_2	b_3	b_4
Douglas-fir	52.8	25	9.015 (0.411)	-0.412 (0.059)	1.0 (0)	0.00091 (0.00049)	1.0 (0)
Mixed conifer	128.0	469	6.579 (0.221)	-0.211 (0.021)	1.0 (0)	0.00110 (0.00007)	1.0 (0)
Ponderosa pine	135.0	59	6.018 (0.357)	-0.002 (0.002)	1.691 (0.131)	0.00306 (0.00061)	1.0 (0)
True fir	81.8	83	7.100 (0.520)	-0.266 (0.054)	1.0 (0)	0.00114 (0.00114)	1.0 (0)

*Standard errors for coefficients are shown parenthetically.

TABLE 11. COEFFICIENTS FOR PREDICTED NUMBER OF UNDERSTORY TREES N_{6-11} * (Model 31)

Timber type	$S_{y.x}$	Number of plots	c_0	c_1	c_2	c_3	c_4
Douglas-fir	35.9	25	6.525 (0.928)	-0.189 (0.122)	0.993 (0.073)	0.00223 (0.00177)	1.0 (0)
Mixed conifer	50.0	468	6.501 (0.219)	-0.146 (0.016)	1.0 (0)	0.00245 (0.00280)	0.870 (0.209)
Ponderosa pine	63.4	56	6.675 (0.832)	-0.188 (0.060)	1.0 (0)	0.00324 (0.00096)	1.0 (0)
True fir	37.5	83	7.252 (0.678)	-0.626 (0.249)	0.836 (0.055)	0.00008 (0.00101)	1.0 (0)

*Standard errors for coefficients are shown parenthetically.

STAG automatically determines the forest type to which the projection set belongs. It uses the classification rules given in table 12.

TABLE 12. CLASSIFICATION RULES FOR FOREST TYPE DETERMINATION

Timber type	Definition
Douglas-fir	Douglas-fir makes up $\geq 80\%$ of the stand basal area (BA_6 or BA_{11})
Ponderosa pine	Ponderosa pine makes up $\geq 80\%$ of the stand basal area (BA_6 or BA_{11})
True fir	Red fir and white fir make up $\geq 80\%$ of the stand basal area (BA_6 or BA_{11})
Mixed conifer	No one species (PP, SP, DF, WF, RE, IC) exceeds 80% of the stand basal area (BA_6 or BA_{11})

It should be noted that these models are accurate but not precise. That is to say, a large variance is associated with these predictions. Standard errors of prediction ($S_{y/x}$) range from 50 to 135 for N_{1-6} and 35 to 65 for N_{6-11} . Therefore, the user is given two options for specifying the number of understory trees. The first option is predicting the number of understory trees by using equations 28 and 29 and either 30 or 31. This predicted number of understory trees for a given stand specification is displayed so that the user can either accept the model prediction or specify another value in lieu of the predicted number. This second option is provided for cases in which the user has good knowledge of local forest conditions and reproduction patterns.

Specification of species

Species of the understory can be specified via two options. In the first option, species composition can be specified to follow the database values used for model development in STAG. These rates are given in table 13.

TABLE 13. PERCENTAGES OF SPECIES BY TIMBER TYPE, ROUNDED TO THE NEAREST 5 PERCENT*

Spp.	Douglas-fir			Mixed conifer			Ponderosa pine			True fir		
	Diameter range			Diameter range			Diameter range			Diameter range		
	1-6	6-11	1-11	1-6	6-11	1-11	1-6	6-11	1-11	1-6	6-11	1-11
PP	5	0	5	10	20	15	50	80	65	0	0	0
SP	5	0	5	5	5	5	5	0	5	5	5	5
IC	10	5	5	30	20	30	15	10	10	10	5	10
DF	60	85	70	15	20	15	5	5	5	0	0	0
WF	20	10	15	40	35	35	25	5	15	75	75	75
RF	0	0	0	0	0	0	0	0	0	10	15	10

*DF₁₋₆ denotes Douglas-fir timber type. The row entries corresponding to this column show the percent of species in the Douglas-fir timber type for trees within the 1- to 6-inch DBH class ($1.5 \leq \text{DBH} \leq 5.5$). Other columns show the percentage of species for a given timber type in the 6- to 11-inch DBH class, and the 1- to 11-inch DBH class. Values of less than 5 percent have been deleted, and the other categories within a column have been proportionally adjusted and rounded to the nearest 5 percent.

We were unable to develop meaningful equations for predicting species composition related to the overstory composition and size of trees. Because of this the percentage of trees occurring in each species in the understory can be specified directly by the user of the program as the second option.

Creating an understory tree list

Because it is possible to generate a large number of understory trees, we use the following methodology to reduce the number of tree records being written. For either the predicted or the user-specified number of understory trees ($0 \leq \text{DBH} \leq \text{minimum [smallest DBH on plot or 11.0 in]}$) we generate individual tree records with a tree expansion factor of 1. The diameters of these trees are generated from the Weibull distribution using the equations for \hat{b} and \hat{c} given

under the heading "Generation of Understory Trees" (eq. 28 and 29). Tree heights and heights-to-crown base are determined according to equations 2 or 3 and 5 or 7, as described under "Estimating Total Height" and "Estimating Height-to-Crown Base," respectively. Stochastic errors are added according to the methods described under "Stochastic Errors."

After the understory is developed in this fashion, the tree records are added to the existing overstory projection set. If the total number of records exceeds the record limit of 500 trees imposed by CACTOS or if the total number of tree records exceeds some user-specified limit (which may be greater or less than 500), the user is given the option of "compressing" the understory tree list. Understory tree record compression is carried out by averaging those tree records that have similar tree attributes, and then replacing those individual tree records with their average values and an appropriate expansion factor.

The compression algorithm is implemented as follows. Individual tree records are grouped into diameter at breast height, total height, and live crown ratio classes by species. If the first grouping does not sufficiently reduce the number of understory tree records, successively coarser and coarser classes are examined until the number of understory tree records is less than or equal to the number desired.

The first grouping uses $\frac{1}{2}$ -inch DBH classes, five dynamically determined height classes, and five dynamically determined live crown ratio classes. The height and live crown ratio classes are dynamically determined, in the sense that the data determine the class limits and class intervals for each live crown ratio class nested within height class, where each height class is nested within a DBH class. Thus, the maximum and minimum heights for the smallest DBH class will generally be different from those in the largest DBH class. Similarly, the largest and smallest live crown ratios found in the smallest height class of the smallest DBH class will generally be different from the largest and smallest live crown ratios found in the largest height class of the smallest DBH class, and so on. We felt that these nested classes would retain more of the "individuality" of each tree record than would nonnested classes.

If the first grouping fails to meet the desired number of understory tree records, the groupings are made successively coarser in the following manner. First, the number of dynamically determined height classes is reduced to three. If this grouping is unsuccessful, the number of live crown ratio classes is reduced from five to three, also. Next, 1-inch DBH classes are tried, then two height classes, then two crown ratio classes, then 2-inch DBH classes. As a last resort, one tree record per understory species is attempted, though a compression of this severity is certainly not recommended if the number of generated understory tree records far exceeds the number of species.

CONVERTING STAND TABLE DATA TO AN INDIVIDUAL TREE LIST

If the user is interested only in the number of trees by diameter class and species, field data can be collected simply by recording the tree species and tallying the DBHs into a diameter class. A table of numbers of trees by diameter class and species produced from this tally is termed a stand table. However, this

form of data collection is not comprehensive because it tells us nothing about the tree heights and crown length or about tree growth. It is a common method for obtaining field data, but obviously individual tree information is lost. Its main advantage is that, since diameters only have to be crudely approximated, there are substantial time savings in collecting information. The number of trees in the diameter classes can be thought of as a discrete approximation to a continuous diameter distribution. Data for producing a stand table can be collected from fixed-area (0.2 or 0.1 acre) or variable-radius (prism) plots. It is generally true that DBHs measured on fixed-area plots provide better approximations to diameter distributions than DBHs measured on prism plots because in the former more trees are sampled. With a variable-radius plot, typically only four to eight sample trees are measured at a sample location, and these trees are selected with probability proportional to size. Hence, larger trees—the trees with greater volume—are frequently selected with variable-radius sampling. Variable-radius plots are efficient for estimating volume but not for representing the total diameter distribution by species of trees in the stand.

In addition, it is common in prism sampling to measure the DBH or DBH class of the sampled trees, but none or few of the sampled trees are actually measured for H or HCB. If missing values are filled in on prism plots, where, say, only DBHs and species have been recorded, the utility of these data for simulation is difficult to assess. Because of these problems, using filled-in data from variable radius plots for simulation in CACTOS may result in less-accurate portrayals of the actual stand conditions, and as a result yield poorer forecasts than those obtained using fixed area sample data.

One way to overcome the disadvantages of variable-radius plots for which few tree attributes are measured is to take enough variable-radius plots to ensure that they provide a good approximation to the true diameter distributions for the species present in the stand. The prism plot stand descriptions should be averaged (using the stand description averager [Meerschaert and Wensel 1987]) and then STAG can be used to fill in missing data for the average stand description or the distributional apportionment routines in STAG can be used to convert diameter class data to pseudo-individual tree data. It is also a good idea to supplement the prism plots with small fixed-area plots (e.g. $\frac{1}{60}$ th or $\frac{1}{100}$ th ac) located the center of each prism plot to better estimate regeneration and small tree frequency.

Because this technique uses a reduced set of field data, it should be more reliable than the stand generation techniques that rely solely on summary statistics to generate the stand description. However, the stand table conversion techniques should be used more cautiously than the missing data routines, since they produce a facsimile of a stand given the reduced data sets provided.

The methodology presented in this section to convert stand table data into individual tree data to produce a facsimile projection set closely parallels the technique used for continuous data that was previously described under the section entitled "Generating Stands from Summary Statistics." We assume that the distribution of grouped diameters, given the species ($p[\text{DBH} \mid \text{species}]$), follows a Weibull distribution. The probability of a tree height falling into some discrete height class, given its species and DBH class ($p[H \mid \text{species}, \text{DBH}]$), and the probability of a tree crown falling into some discrete class, given its species, DBH, and height class ($p[\text{HCB} \mid \text{species}, \text{DBH}, H]$), are both hypothesized to

follow a normal distribution. These assumptions were tested using a Kolmogorov-Smirnov test and found to be acceptable. For further detail, see Van Deusen (1984).

Diameter Distributions

We postulated that the distribution of diameters (or diameter classes) followed a Weibull distribution, but within a given diameter class we assumed that numbers of trees followed a uniform distribution. If diameter classes are not wide, then this is a plausible assumption. We tested these assumptions on 50 $\frac{1}{4}$ -acre plots (see Van Deusen 1984) and found that the simplifying assumption of uniform distribution of diameters within a diameter class yielded results quite similar to that obtained by using a Weibull distribution for DBH classes when diameter classes were no larger than 2 inches.

Height Distributions

An average value for height of overstory trees is predicted from equations 1 and 4, or for understory tree height from models 2 or 3, by using the diameter class mean value for DBH. The predicted average height is used to locate the centroid of the height distribution, which is assumed to follow a normal distribution within a diameter class (see figure 2). The variance of the distribution is then approximated using the variance of the regression of the height prediction equation. We estimate the proportion of trees to allocate to a specific height class within a diameter class by determining the percentage of the area under the curve for each height class. We call this process distributional apportionment because we allocate (apportion) the number of trees per diameter class over the height classes using this methodology.

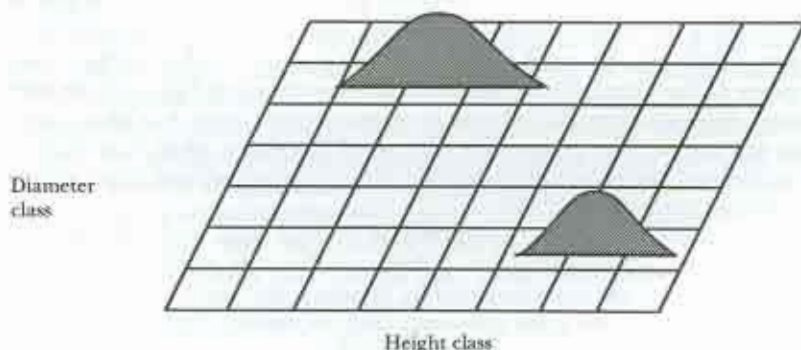


Fig. 2. Distributional apportionment of stand table data.

Height-to-Crown Base Distribution

We assume that the distribution of heights-to-crown base within a given height and DBH class follows a normal distribution. We allocate the numbers of trees into each of the crown classes with the same methodology used for allocating trees into height classes. The normal curve is first located using the mean height-to-crown base value from equation 4, or models 5 and 7 for understory tree

height-to-crown base, assuming the midpoints of the height and diameter class. The variance of the normal curve is approximated by the variance about the height-to-crown base predictive model. In the last step, the area under the curve within each crown class is calculated and the number in each height-diameter cell within the crown class is determined by multiplying these proportions by the number of trees in a given height-diameter class.

This apportioning process calculates the numbers of trees to place in each cell of the height-diameter-crown categories. We define these cells to be either 1- or 2-inch-diameter classes (specified by the user), 10-foot height classes, and 10-foot height-to-crown base classes. Individual tree dimensions (diameter, height, and height-to-crown base) are given an equal probability of occurring at any location within this three-dimensional cell by drawing random numbers that correspond to x, y, z coordinates in 3-D space. Using this procedure we have developed an individual tree list from the original stand table, but they are pseudo-individual, in the sense that they have been estimated using the above procedure rather than measured.

VALIDATING THE DIAMETER DISTRIBUTION GENERATION PROCEDURE

The procedure was tested on 166 $\frac{1}{4}$ -acre permanent plots from the Northern California Forest Yield Cooperative database. Only the Weibull distribution was tested, since the stands used for the test are considered generally to be managed stands and do not usually follow a negative exponential distribution. The second measurement of these data occurred in 1984. For the test we used plots from the southern Cascade region. The accuracy of the procedure for predicting the number of trees per DBH class and the volume per DBH class was evaluated. We used 2-inch DBH classes, beginning at 5.5 inches and going to 49.5 inches DBH. The error index developed by Reynolds, Burk, and Huang (1988) was used for the test and is given as

$$e = N \times \sum_{j=1}^k \left| \int_{I_j} w(x) \times d\hat{F}(x) - \int_{I_j} w(x) \times dF(x) \right| \quad [32]$$

- where e = error index
 N = number of trees per acre
 $w(x)$ = weighting factor
 $\hat{F}(x)$ = the cdf of diameters on a plot as predicted from the model
 $F^*(x)$ = the empirical cdf
 $dF(x)$ = the differential of the cdf (empirical or predicted) with respect to x (diameter)
 k = the number of DBH classes
 I_j = the j^{th} DBH class.

As the authors of the index point out, a good fit in one diameter class does not offset a poor fit in another. The error index provides a means for comparing

the overall fit of a model to another model, but the individual cells (DBH, species classes) must be examined to determine where a particular model fits adequately.

We performed two sets of analysis. In the first we computed the error index of an "average stand." The average stand is the stand table produced by averaging all of the 166 stand tables associated with each of these plots. We judged our ability to produce a tree diameter distribution by seeing how accurately the number of trees in various diameter classes is predicted for this average stand. We also judged how well our diameter distribution models work by comparing the volumes (Biging 1983) predicted for each diameter class with the average volume computed from the 166 test plots. In the second analysis we presented results that show the average of the error indices computed for each plot individually.

The results for the average stand are shown in tables 14 and 15. Table 14 shows the "misclassification" by species and DBH class for the average of the 166 plots. By misclassification we mean the signed values calculated from differencing the predicted number of trees (or volumes) from the actual number of trees (or volumes) in each diameter class. The sum of the absolute values (predicted minus observed, see equation 32) is used by Reynolds, Burk, and Huang (1988) to calculate the error index. Thus, for ponderosa pine in the 8.5-inch DBH class, this model underpredicted by two trees (see table 14).

In the right margin of table 14 the error indices are listed by species. The indices' magnitude corresponds, relatively, to the abundance of the trees on the plots. In other words, the more trees there are, the greater the error. The bottom margin is the average misclassification across species for a particular DBH class. Thus, we see on average an underprediction for the 6.5- to 8.5-inch DBH classes and an overprediction in the 10.5- to 12.5-inch DBH classes. There are on average only slight underpredictions for the 18.5- to 26.5-inch DBH classes. The lower right cell of table 14 provides the overall error index for this "averaged" plot, which is a value of 68.

Another statistic we computed was the average plot error index with its associated standard errors. The average was 358 and the standard error was 11.8. The average is quite large and shows the difficulty of predicting the diameter distribution for a particular plot. The error index in table 14 is much smaller (68) because we are averaging the plots and then computing the errors, as opposed to an average error index value (358.13) calculated as the average of the individual plot error indices.

Table 15 provides the same type of information as table 14. In table 15 the error index is weighted by board foot volume, whereas in table 14 the error index is weighted by numbers of trees. In table 15 we see that volumes are on average slightly underpredicted for the 6.5- to 8.5-inch diameter class. Volumes are overpredicted in the 10.5- to 18.5-inch diameter class, underpredicted in the 20.5- to 28.5-inch diameter class, and overpredicted in the 30.5-inch and 32.5-inch and greater diameter classes.

In table 14 we reported that on average the models underpredicted by two or three trees in the 18.5- to 26.5-inch diameter class. Because trees in this size range average around 200 to 500 board feet, it is not surprising that in table 15 we find that the misclassification index for diameter classes in this range varies

from an overprediction of 516 board feet to an underprediction of 2,374 board feet. The average net effect of these over- and underpredictions is a slight overprediction of 330 board feet. Thus, there appears to be no major bias in volume or numbers of trees associated with producing diameter distributions using the Weibull generation procedure.

We also computed the average plot error index weighted by volume with its associated standard errors. The average was 44,540 board feet and the standard error was 3,187 board feet. Again, this underscores the difficulty of accurately predicting the diameter and volume distribution on any particular plot.

In another test of this procedure we used the same plots to create "known"

TABLE 14. AVERAGE MISCLASSIFICATION INDICES (ACTUAL MINUS PREDICTED) OF NUMBERS OF TREES PER ACRE BY SPECIES AND DBH CLASSES AND OVERALL ERROR INDEX OF THE "AVERAGE" PLOT FROM STAG VERSION 4.2*

Misclassification Index																	
Species	DBH class															Species	
	6.5	8.5	10.5	12.5	14.5	16.5	18.5	20.5	22.5	24.5	26.5	28.5	30.5	≥32.5	error	index	
PP	3	2	-4	-6	0	0	0	1	1	0	1	0	0	0	-2	(18)	
SP	1	0	0	0	1	0	0	0	0	0	0	0	0	0	2	(2)	
IC	7	3	0	0	0	0	0	0	0	0	0	0	0	0	12	(12)	
DF	5	0	-2	0	0	0	0	0	1	0	0	0	0	0	4	(8)	
WF	9	4	-5	-5	-1	0	2	1	1	1	0	0	0	0	7	(29)	
RF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	(0)	
	25	11	-11	-11	0	0	2	2	3	1	1	0	0	0	23	(68)	

*The value in the penultimate column is the signed value, while next to it in parenthesis is the absolute value necessary for the computation of the error index.

TABLE 15. AVERAGE MISCLASSIFICATION INDICES (ACTUAL MINUS PREDICTED) OF BOARD FEET TO A 6-INCH TOP PER ACRE BY SPECIES AND DBH CLASSES FROM STAG VERSION 4.2*

Misclassification Index																
Species	DBH class														Species	
	6.5	8.5	10.5	12.5	14.5	16.5	18.5	20.5	22.5	24.5	26.5	28.5	30.5	≥32.5	error	index
PP	38	24	-221	-521	-345	-223	-295	318	671	186	983	-95	-288	-576	-344	(5,394)
SP	2	-11	-6	-6	20	58	1	-110	60	125	-28	61	-67	-440	-341	(1,651)
IC	7	-1	-51	-72	-35	-130	87	4	130	90	18	0	72	269	388	(988)
DF	21	-27	-193	-208	-215	-216	-299	-199	762	-3	199	0	-23	-161	-562	(2,548)
WF	84	55	-369	-716	-652	-624	42	309	655	205	107	802	-38	526	386	(5,236)
RF	3	-1	-44	-54	-57	-27	-52	18	96	72	384	-22	-69	-104	143	(1,003)
	155	39	-884	-1,577	-1,284	-1,162	-516	340	2,374	675	1,663	746	-415	-486	-330	(16,820)

*The value in the penultimate column is the signed value, while next to it in parenthesis is the absolute value necessary for the computation of the error index.

TABLE 16. AVERAGE STAND TABLE BASED ON 166 PERMANENT PLOTS FROM THE SOUTHERN CASCADE REGION

Diameter class	Observed av. numbers	Expected av. numbers	Observed av. height	Expected av. height	Observed av. HCB	Expected av. HCB
6.5	50.2	21.7	35.5	39.7	21.0	16.3
8.5	49.7	35.9	45.2	48.4	26.1	21.3
10.5	30.4	42.5	56.4	57.6	31.9	26.3
12.5	28.9	40.7	65.0	65.5	36.3	30.5
14.5	29.5	29.7	72.2	73.4	38.5	34.6
16.5	18.8	19.8	80.8	81.0	42.7	38.3
18.5	15.2	12.5	85.7	87.9	44.7	42.4
20.5	10.3	8.0	95.0	95.0	49.0	42.5
22.5	9.8	4.5	99.1	102.3	51.6	49.0
24.5	4.6	3.0	106.6	109.2	52.7	51.8
26.5	5.0	1.9	113.1	114.8	59.4	55.2
28.5	3.0	1.5	116.3	120.0	61.3	58.1
30.5	1.3	1.0	120.0	124.8	65.8	60.7
32.5	0.6	0.5	119.6	132.1	61.7	65.8
34.5	0.7	0.4	135.2	136.6	75.5	69.9
36.5	0.6	0.3	121.9	149.3	64.6	71.3
38.5	0.4	0.2	140.4	145.8	69.6	68.5
40.5	0.1	0.1	135.2	157.7	76.6	73.6
42.5	0.1	0.1	147.7	168.8	89.7	78.1
44.5	0.0	0.0	—	170.3	—	78.1
46.5	0.1	0.1	146.3	167.8	97.0	83.1
48.5	0.0	0.0	—	169.8	—	85.3

stand tables. We then used the stand tables to apportion the trees over height and crown classes. The results for the average stand table based on these 166 plots are presented in table 16. The numbers of trees apportioned into these classes corresponded well with the actual numbers observed on the plots, except for the smallest diameter classes. Predicted heights and predicted heights-to-crown base were generally close to the observed average values. This demonstrates that stand tables can be generated that on average closely approximate actual stands. Good judgment should be exercised in using these routines. Real field data is always preferable to generating stands from summary statistics. Even though these procedures produce reasonable facsimiles to real stands, this process always produces inaccuracies. For a more detailed treatment of this analysis, see Van Deusen (1984).

DISCUSSION AND RECOMMENDED USES OF STAG

The Forest Stand Generator, STAG, is an important component of a simulation system for mixed conifer growth and yield projection. STAG was created to ensure that different types of inventory data could be supplemented to produce data sets suitable for projection in the forest simulator CACTOS. There are different procedures and analysis routines within STAG for (1) generating missing data, (2) converting stand table data (approximations to a diameter distribution), and (3) transforming summary statistics, such as number of trees and basal area per

acre, to a projection set composed of complete individual tree records for use in CACTOS. To fill in missing data, STAG uses predictive equations for total height (H) and height-to-crown base (HCB) developed from a permanent plot system of over 20,000 trees in northern California. To create a projection set based only on summary statistics (termed *stand generation*) is much more complicated. To accomplish this, STAG factors the joint distribution for species, DBH, H, and HCB into a product of probability density functions, and models each of these components. The methodology developed for converting stand table data closely follows that described for stand generation.

We developed these procedures to increase the availability of data that can be used with the CACTOS simulation system, not to encourage a shift away from field data collection. The best and most reliable use of STAG is in combination with statistically valid field data plots (usually measuring DBH, H, HCB, and species of trees on 0.2- or 0.1-acre field plots) to fill in occasional missing data values. Even though we have developed stand generation techniques for producing facsimiles of stands from summary statistics, users of this technique should be aware that this may produce highly variable results. We have done limited testing of the stand generation procedures using permanent plot data for mixed-species, multiple-aged coniferous stands, and have found, in these test cases, that the stand generation produced reasonable facsimiles of stands. Because we tested only a small subset of the possible types of stands that could be generated, we cannot say that this technique can be generally applied with good results. As such, users should use the stand generation techniques with great caution. We recommend that stand generation be used only as a last resort, not as a matter of course. Remember that there is no replacement for real field data. Converting stand table data to a projection set is a case that is intermediate between filling in missing data and stand generation. Stand table data are counts of the number of trees by species observed on field plots that fall within specified diameter classes. These field data can be collected simply by recording the tree species and tallying the DBH into a diameter class. However, this form of data collection is not comprehensive, because it tells us nothing about the tree heights and crown lengths. This frequency information can be used to approximate a continuous diameter distribution, and heights and heights-to-crown bases can be generated to complete the stand description. Because this technique uses a reduced set of field data, it should be more reliable than the stand generation techniques, which rely solely on summary statistics to generate the stand description. However, the stand table conversion techniques should be used more cautiously than the missing data routines, since they produce a facsimile of a stand given the reduced data sets provided.

APPENDIX A.
SUMMARY STATISTICS FOR PERMANENT PLOT TREE DATA

Ponderosa pine n = 4,173 trees					
Variable	Median	Mean	Std. deviation	Minimum	Maximum
DBH (in)	12.9	14.12	6.58	5.5	55.8
Total height (ft)	71.0	72.87	27.97	12.0	184.0
Height-to-crown base (ft)	32.0	34.92	18.21	1.0	145.0
Site index (ht @ 50 yrs)	75.0	74.20	17.51	29.0	150.0
Elevation (ft)	4,100.0	4,150.00	786.40	2,150.0	7,500.0
Basal area (ft ²) per acre	189.6	194.42	82.31	22.5	532.7
Number of trees per acre	200.0	215.43	100.04	16.0	515.0
Sugar pine n = 1,070 trees					
Variable	Median	Mean	Std. deviation	Minimum	Maximum
DBH (in)	14.1	15.96	8.41	5.5	59.1
Total height (ft)	71.0	73.49	31.46	15.0	199.0
Height-to-crown base (ft)	34.0	36.91	18.25	1.0	105.0
Site index (ht @ 50 yrs)	78.0	76.67	16.30	29.0	150.0
Elevation (ft)	4,640.0	4,566.00	733.80	2,350.0	7,000.0
Basal area (ft ²) per acre	197.6	216.27	98.08	30.2	532.7
Number of trees per acre	188.0	207.89	90.64	15.0	490.0
Incense cedar n = 2,260 trees					
Variable	Median	Mean	Std. deviation	Minimum	Maximum
DBH (in)	11.0	12.70	6.90	5.5	67.6
Total height (ft)	43.0	48.19	22.17	11.0	182.0
Height-to-crown base (ft)	22.0	25.10	14.88	1.0	95.0
Site index (ht @ 50 yrs)	76.0	76.22	16.49	29.0	130.0
Elevation (ft)	4,420.0	4,443.00	735.90	2,150.0	6,640.0
Basal area (ft ²) per acre	203.0	213.25	90.05	27.2	532.7
Number of trees per acre	192.0	205.21	91.31	18.0	515.0
Douglas-fir n = 2,458 trees					
Variable	Median	Mean	Std. deviation	Minimum	Maximum
DBH (in)	12.2	13.22	5.91	5.5	50.4
Total height (ft)	71.0	72.59	24.92	11.0	170.0
Height-to-crown base (ft)	36.0	38.63	18.34	2.0	126.0
Site index (ht @ 50 yrs)	77.0	77.82	16.59	36.0	157.0
Elevation (ft)	3,700.0	3,802.00	784.00	2,150.0	5,880.0
Basal area (ft ²) per acre	163.6	169.06	72.56	16.5	424.2
Number of trees per acre	170.0	182.98	76.14	20.0	515.0
White fir n = 5,167 trees					
Variable	Median	Mean	Std. deviation	Minimum	Maximum
DBH (in)	11.7	13.03	6.17	5.5	48.9
Total height (ft)	61.0	64.29	26.60	9.0	171.0
Height-to-crown base (ft)	31.0	33.37	17.49	1.0	114.0
Site index (ht @ 50 yrs)	76.0	76.29	16.50	23.0	130.0
Elevation (ft)	4,850.0	4,862.00	752.40	2,700.0	7,500.0
Basal area (ft ²) per acre	204.0	211.82	89.91	3.5	532.7
Number of trees per acre	188.0	203.20	89.92	5.0	525.0

(Continued on next page)

APPENDIX A. (Continued)

Red fir n = 501 trees					
Variable	Median	Mean	Std. deviation	Minimum	Maximum
DBH (in)	13.8	15.39	7.54	5.5	51.6
Total height (ft)	69.0	70.01	29.09	15.0	154.0
Height-to-crown base (ft)	32.0	33.55	18.63	3.0	92.0
Site index (ht @ 50 yrs)	63.0	65.43	11.35	46.0	104.0
Elevation (ft)	5,800.0	5,840.00	547.30	4,640.0	7,500.0
Basal area (ft ²) per acre	213.7	230.00	94.37	36.6	428.8
Number of trees per acre	176.0	199.41	95.01	15.0	525.0

Other hardwoods n = 273 trees					
Variable	Median	Mean	Std. deviation	Minimum	Maximum
DBH (in)	10.0	10.79	4.41	5.5	25.2
Total height (ft)	44.0	47.57	19.85	11.0	104.0
Height-to-crown base (ft)	24.0	25.36	11.62	2.0	69.0
Site index (ht @ 50 yrs)	82.0	78.51	21.52	37.0	114.0
Elevation (ft)	3,560.0	3,779.00	736.20	2,700.0	5,500.0
Basal area (ft ²) per acre	183.2	204.61	108.70	22.5	454.4
Number of trees per acre	160.0	173.97	65.70	36.0	305.0

Black oak n = 340 trees					
Variable	Median	Mean	Std. deviation	Minimum	Maximum
DBH (in)	10.8	12.93	7.00	5.5	52.7
Total height (ft)	51.5	52.55	19.55	12.0	164.0
Height-to-crown base (ft)	22.0	24.75	13.92	1.0	82.0
Site index (ht @ 50 yrs)	75.0	75.83	15.45	37.0	114.0
Elevation (ft)	4,218.0	4,116.00	806.70	2,350.0	5,700.0
Basal area (ft ²) per acre	172.65	188.40	86.57	30.6	424.2
Number of trees per acre	168.0	184.35	79.71	16.0	376.0

APPENDIX B.
SUMMARY STATISTICS FOR PERMANENT PLOT SMALL TREE DATA

All species Tree statistics n = 7,157 trees					
Variable	Mean	Std. deviation	Minimum	Maximum	Median
DBH (in)	3.31	1.12	1.50	5.40	3.20
Total height (ft)	18.66	8.00	5.00	64.00	17.00
Height-to-crown base (ft)	10.00	6.55	1.00	56.00	8.00

Plot statistics n = 308 plots					
Basal area of all trees $\leq 5.4"$	0.84	8.78	0.02	9.32	0.43
Number of all trees $\leq 5.4"$	176.18	183.00	4.00	1,400.00	120.00
Elevation (ft)	2,150	919.3	4,519	7,300	4,510

LITERATURE CITED

- AVERY, T. E., and H. E. BURKHART
 1983. *Forest measurements*. New York: McGraw-Hill.
- BIGING, G. S.
 1983. *Volume tables for young-growth mixed conifers of northern California based upon the stem analysis data. Part I; Volume tables for mixed conifers. Part II; Taper equations for mixed conifers*. Research Note 7, revised draft. University of California, Berkeley: Northern California Forest Yield Cooperative, Department of Forestry and Resource Management.
- BIGING, G. S.
 1984. Taper equations for second-growth mixed conifers of northern California. *For. Sci.* 30(4):1103-17.
- BIGING, G. S. and L. C. WENSEL
 1987. STAG: A forest STAND Generator for producing complete CACTOS stand descriptions. In A. Ek and T. Burk, eds., *Forest Growth Modelling and Prediction*, Vol. 1, 47-53. USDA Forest Service General Technical Report NC-120.
- BIGING, G. S., E. C. TURNBLOM, W. J. MEERSCHAERT, and T. A. ROBARDS
 1995. *STAG user's guide: The Forest Stand Generator for Mixed Conifer Species in California*. Version 4.5. Manuscript for the University of California Agricultural Experiment Station Bulletin series.
- BURK, T. E., M. H. HANSEN, and A. R. EK
 1982. *Combining sources of information for improved in-place inventory statistics*. Orno, Me: In Place Inventories.
- DAVIS, L. S., and K. N. JOHNSON
 1987. *Forest Management*, 3rd ed. New York: McGraw-Hill.
- EK, A. R.
 1974. Nonlinear models for stand table projection in northern hardwood stands. *Can. J. For. Res.* 4:23-27.
- HUSCH, B., C. I. MILLER, and T. W. BEERS
 1982. *Forest Mensuration*, 3rd ed. New York: John Wiley & Sons.
- HYINK, D. M., and J. W. MOSER
 1983. A generalized framework for projecting forest yield and stand structure using diameter distributions. *For. Sci.* 29:85-95.
- KNOEBEL, B. R., and H. E. BURKHART
 1991. A bivariate distribution approach to modeling forest diameter distributions at two points in time. *Biometrics* 47(1):241-53.
- KNOX, R. G., R. K. PEET, and N. L. CHRISTENSEN
 1989. Population dynamics in loblolly pine stands: Changes in skewness and size inequality. *Ecology* 70:1153-66.
- MEERSCHAERT, W. J., and L. C. WENSEL
 1987. *SDAVG User's Guide: The CACTOS Stand Description Averager Version 2.0*. Research Note 19, Northern California Forest Yield Cooperative. Berkeley: University of California.
- MEYER, H. A.
 1952. Structure, growth, and drain in balanced uneven-aged forests. *J. For.* 50(2):85-92.
- MILLER, T. E., and J. WEINER
 1989. Local density variation may mimic effects of asymmetric competition on plant size variability. *Ecology* 70:1188-91.
- REINEKE, L. H.
 1933. Perfecting a stand density index for even-aged forest. *J. Agric. Res.* 46:627-38.
- REYNOLDS, M. R., T. E. BURK, and W. C. HUANG
 1988. Goodness of fit tests and model selection procedures for diameter distribution models. *For. Sci.* 34(2):373-99.
- SHIFLEY, S. R.
 1990. Analysis and modeling of patterns of forest ingrowth in the North Central United States. Ph.D. dissertation, St. Paul, Minn.: University of Minnesota.
- SHIFLEY, S. R., and E. LENTZ
 1985. Quick estimation of the three-parameter Weibull to describe tree size distributions. *Forest Ecology and Mgt.* 13:195-203.
- VAN DEUSEN, P. C.
 1984. Complete and partial generation of tree characteristics for mixed species stands. Ph.D. dissertation, Berkeley: University of California.
- WENSEL, L. C., and G. S. BIGING
 1987. The CACTOS system for individual-tree growth simulation in the mixed conifer forests of California. In A. Ek and T. Burk, eds., *Forest Growth Modelling and Prediction*, Vol. 1, 175-83. USDA Forest Service General Technical Report, NC-120.
- WENSEL, L. C., P. J. DAUGHERTY, and W. J. MEERSCHAERT
 1986. CACTOS user's guide: The California conifer timber output simulator. Berkeley: University of California, Division of Agricultural Science Bulletin 1920.
- WENSEL, L. C., W. J. MEERSCHAERT, and G. S. BIGING
 1987. Tree height and diameter growth models for northern California conifers. *Hilgardia* 55(8):1-20.

Continued from inside front cover

Research reported here was conducted under UC Agricultural Experiment Station projects 3679-ms and 3815-ms. The authors gratefully acknowledge important contributions to this work by Peter J. Daugherty and Vaughan Landrum. They also are grateful to Lee C. Wensel and the members of the Northern California Forest Yield Cooperative for their support.

READER PLEASE NOTE:

This issue of *Hilgardia* begins Volume 61. The previous issue of *Hilgardia* (Volume 60, Number 1) was the only number published in Volume 60.

The University of California, in accordance with applicable Federal and State law and University policy, does not discriminate on the basis of race, color, national origin, religion, sex, disability, age, medical condition (cancer-related), ancestry, marital status, citizenship, sexual orientation, or status as a Vietnam-era veteran or special disabled veteran. The University also prohibits sexual harassment. Inquiries regarding the University's nondiscrimination policies may be directed to the Affirmative Action Director, University of California, Agriculture and Natural Resources, 300 Lakeside Drive, 6th Floor, Oakland, CA 94612-3560, (510) 987-0096.

Issued in furtherance of Cooperative Extension work, Acts of May 8 and June 30, 1914, in cooperation with the U.S. Department of Agriculture. Kenneth R. Fairrell, Director of Cooperative Extension, University of California.

HILGARDIA Editorial Board

Edward S. Sylvester, Chair, Berkeley
(entomology, insecticides, ecology, environmental toxicology)

Peter Berck, Associate Editor, Berkeley
(economics, statistics, resource management)

Harry W. Colvin, Associate Editor, Davis
(animal science, physiology, breeding, zoology, genetics)

Donald J. Durzan, Associate Editor, Davis
(tree fruit and nut crops)

Walter G. Jennings, Associate Editor, Davis
(food science, nutrition, chemistry)

John Letey, Associate Editor, Riverside
(soils, plant nutrition, agronomy, agricultural engineering, water)

Irwin P. Ting, Associate Editor, Riverside
(botany, plant physiology, biochemistry)

Janet White, Managing Editor, Oakland

The Journal HILGARDIA is published irregularly. Number of pages and number of issues vary per annually numbered volume. Address: Agriculture and Natural Resources Publications, University of California, 300 Lakeside Drive, 6th Floor, Oakland, CA 94612-3550.